Incremental Distributed Robust Inference from Arbitrary Robot Poses via EM and Model Selection

Vadim Indelman*^, Nathan Michael†, and Frank Dellaert*

***Abstract*—We present a novel approach for multi-robot distributed and incremental inference over variables of interest, such as robot trajectories, considering the initial relative poses between the robots and multi-robot data association are both unknown. Assuming robots share with each other informative observations, this inference problem is formulated within an Expectation-Maximization (EM) optimization, performed by each robot separately, alternating between inference over variables of interest and multi-robot data association. To facilitate this process, a common reference frame between the robots should first be established. We show the latter is coupled with determining multi-robot data association, and therefore concurrently infer both using a separate EM optimization. This optimization is performed by each robot starting from several promising initial solutions, converging to locally-optimal hypotheses regarding data association and reference frame transformation. Choosing the best hypothesis in an *incremental* problem setting is in particular challenging due to high sensitivity to *perceptual aliasing* and possibly insufficient amount of data. Selecting an incorrect hypothesis introduces outliers and can lead to catastrophic results. To address these challenges we develop a model-selection based approach to choose the most probable hypothesis and use the Chinese restaurant process to disambiguate the hypotheses prior probabilities over time.**

I. Introduction

Distributed inference is a key capability in multi-robot autonomous systems that is of interest in a variety of problem domains, including surveillance, tracking, localization and mapping. Cooperatively inferring variables of interest, such as robot trajectories, observed objects and tracked targets, results in higher levels of performance, flexibility and robustness to failure. The research community has been addressing different aspects of this problem, considering both centralized and decentralized frameworks (e.g. [10, 3, 6, 4]).

To facilitate cooperative inference it is essential to establish a common reference frame and world model between the robots, so that these can communicate with each other relevant information and correctly interpret it for their needs. While each of these problems has been previously addressed assuming the other problem is solved, only few attempts have been made to solve the two problems *simultaneously*: determining a common reference frame between the robots, and resolving data association between measurements (e.g. images or laser scans) acquired by different robots.

Solving these coupled problems is important as it enables the robots, scattered in a previously unknown environment, to establish collaboration without requiring any prior knowledge or infrastructure. For example, starting from arbitrary guess

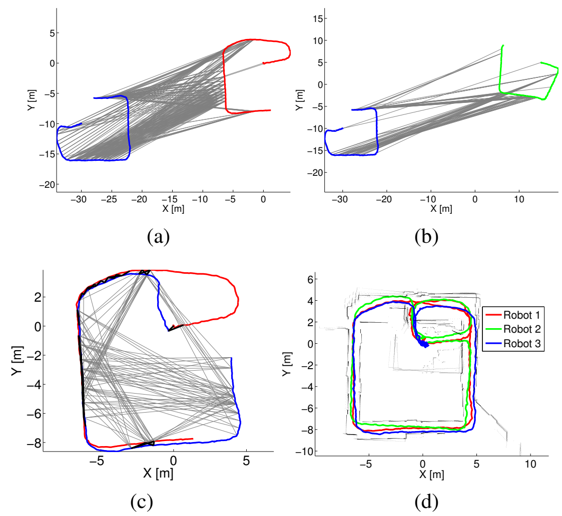


Figure 1: Multi-robot correspondences $\mathcal{F}^r$ between (a) red and blue, and (b) green and blue robots. As a common reference frame is not yet established, robot initial poses are set to arbitrary values.(c) Blue robot estimates of its own and red robot's trajectories expressed right after a common reference frame between the two robots has been established. The identified inlier (black) and outlier (gray) correspondences right. (d) Ground truth.

as to where each robot is and by sharing measurements of onboard sensors, each robot will be capable of inferring the trajectories of other robots in the group (Fig. 1).

Multi-robot data association is a key challenge that shares some similarities with loop closure detection in the single-robot case. Incorrect data association can lead to catastrophic deterioration in performance and should be avoided at all costs; the robotics community has been indeed very active in the last two years in developing robust graph optimization techniques [7, 11, 8] to address this crucial aspect.

Multi-robot data association has recently become an active research area as well (e.g. [2, 9, 5]), with the same sensitivity to incorrect correspondences as in the single-robot case. This problem, however, becomes more complicated when the initial relative poses between the robots are unknown. Without a common reference frame, how can the robot decide what information to share with each other? Given the calculated multi-robot constraints based on this shared information, how to determine the inlier correspondences? Addressing this problem requires reasoning about multi-robot data association and initial relative poses concurrently.

Recently, an Expectation-Maximization (EM) approach has been developed in [8] for the single-robot case, as well as by the authors [5] in the multi-robot problem setting. Here, we

* Institute for Robotics and Intelligent Machines (IRIM), Georgia Institute of Technology, Atlanta, GA 30332, USA.

^ Faculty of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel.

†Robotics Institute, Carnegie Mellon, Pittsburgh, PA 15213, USA.

present a distributed and incremental multi-robot approach that allows a group of robots to simultaneously establish a common reference frame and resolve multi-robot data association on-the-fly. To that end, each robot performs an EM optimization from several initial guesses, resulting in different locally-optimal solutions.

Our approach is based on the observation [5] that by analyzing the distribution of multi-robot relative pose constraints (Figs. 3a and 4a) it is possible to estimate the transformation between the robot reference frames and identify the inliers in these constraints. Choosing the correct solution is a key challenge, as a wrong decision will effectively introduce outliers to the graph optimization. This is particularly true when information is received *incrementally*: in this case, one needs also to decide whether *sufficient* amount of information has been received to perform this decision reliably.

Furthermore, as we discuss in the sequel, perceptual aliasing presents additional challenges, leading to additional *clusters* that compete with the inliers cluster, and the problem becomes how to choose the right cluster among several candidates.

We consider this challenging problem and frame it within a *model selection* framework, developing a probabilistically-sound approach for selecting the most probable cluster. Moreover, we address the question wether there *is* a correct cluster given the information available to each robot thus far, as the robots might have not observed the same environment yet. We approach this problem by modeling the prior probability for each cluster using the *Chinese restaurant process* (e.g., [1]) that allows to disambiguate this decision-making as more information is accumulated.

II. Problem Formulation

We consider a group of R robots deployed to collaboratively operate in an unknown environment, *initially unaware* of each other. Our objective is for each robot r to estimate $\mathcal{X}^r$ which comprises its own trajectory X^r (current and past poses) and additional variables of interest, such as the trajectories of other robots, in a distributed incremental framework. In a distributed setting the inference solved by each robot r is

$$\hat{\mathcal{X}}^r = \underset{\mathcal{X}^r}{\arg\max}\, p\left(\mathcal{X}^r|\mathcal{Z}^r\right), \tag{1}$$

where $\mathcal{Z}^r$ represents the available measurements to robot r: its own observations and observations shared by other robots.

We assume each robot r shares at each time t_k its current measurement z_k^r, if it is informative, and also tracks all these informative measurements $\{z_i^r\}$ over time. Any robot r that receives a measurement $z_k^{r'}$ from some robot r', generates candidate correspondences by matching $z_k^{r'}$ with its own informative measurements $\{z_i^r\}$. Each such successful match $u_{k,l}^{r',r}$ between $z_k^{r'}$ and $z_l^r \in \{z_i^r\}$ represents a (relative-pose) constraint involving the poses $x_k^{r'}$ and x_l^r, with $l \leq k$. We denote by $\mathcal{F}^r$ the set of multi-robot data association, that is available to robot r, where each individual data association $(r',k,l) \in \mathcal{F}^r$ represents the constraint $u_{k,l}^{r',r}$.

An example of the multi-robot candidate correspondences set $\mathcal{F}^r$ is shown in Fig. 1. The figure illustrates the candidate correspondences in $\mathcal{F}^r$ between the blue robot and other robots (green and red). Since the initial relative poses between the robots are unknown, these transformations were set to *arbitrary* values, i.e. the initial pose of each robot was chosen arbitrarily.

Observe that many of these correspondences in $\mathcal{F}^r$ are outliers. Moreover, *perceptual aliasing* will often result in *numerous* false data associations that are *consistent* with each other: A typical example is matching between laser scans (or images) from different but similar in appearance corridors; the result of this match will typically erroneously indicate the two places are the same. Not only this estimate is completely wrong, but also *similar* erroneous estimates will be obtained for all such matches, making it difficult to identify these are all outliers.

III. Approach

Given the multi-robot data association $\mathcal{F}^r$, and the appropriate constraints $u_{k,l}^{r',r}$, the joint pdf from Eq. (1) can be expressed as

$$p(\mathcal{X}^r|\mathcal{Z}^r) \propto p\left(X^r|Z^r\right) p\left(\mathcal{X}^{\mathcal{R}\setminus\{r\}}|\mathcal{Z}_{local}^{\mathcal{R}\setminus\{r\}}\right) \prod_{(r',k,l)\in\mathcal{F}^r} p\left(u_{k,l}^{r',r}|x_k^{r'},x_l^r\right), \tag{2}$$

where the set $\mathcal{X}^{\mathcal{R}\setminus\{r\}}$ represents all the poses $x_i^{r'}$ of robots $r' \in \mathcal{R}$ that contributed at least one correspondence to $\mathcal{F}^r$, and $\mathcal{Z}_{local}^{\mathcal{R}\setminus\{r\}}$ are the local observations of these robots.

As the robots express their local trajectories with respect to *different* reference systems, the measurement likelihood term in Eq. (2) is

$$p\left(u_{k,l}^{r',r}|x_k^{r'},x_l^r\right) \propto \exp\left(-\frac{1}{2}\left\|err\left(u_{k,l}^{r',r},x_k^{r'},x_l^r\right)\right\|_\Sigma^2\right), \tag{3}$$

with

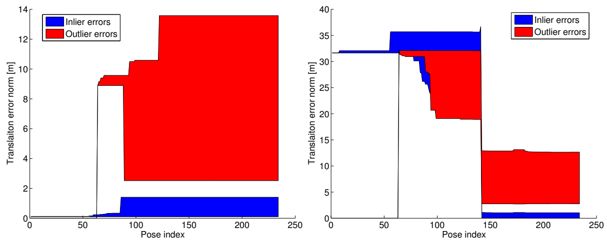
$$err\left(u_{k,l}^{r',r},x_k^{r'},x_l^r\right) \doteq u_{k,l}^{r',r} \ominus h\left(x_k^{r'},x_l^r\right), \tag{4}$$

and $h\left(x_k^{r'},x_l^r\right) \doteq x_k^{r'} \ominus \left(T_r^{r'} \oplus x_l^r\right)$. The transformation $T_r^{r'}$ represents a common reference frame between robots r and r'. This transformation, as well as multi-robot data association are *both* unknown.

Instead of assuming multi-robot data association is given, we introduce a latent binary variable $j_{k,l}^{r',r}$ for each correspondence $(r',k,l) \in \mathcal{F}^r$ and use the convention that this correspondence is an inlier if $j_{k,l}^{r',r} = inlier$ and accordingly outlier when $j_{k,l}^{r',r} = outlier$. Denoting all the latent variables representing data association between robot r and other robots by $\mathcal{J}^r$ and considering it to be part of the inference, the probabilistic formulation (2) turns into:

$$p\left(\mathcal{X}^r,\mathcal{J}^r|\mathcal{Z}^r\right) \propto p\left(X^r|Z^r\right) \cdot p\left(\mathcal{X}^{\mathcal{R}\setminus\{r\}}|\mathcal{Z}_{local}^{\mathcal{R}\setminus\{r\}}\right) \prod_{(r',k,l)\in\mathcal{F}^r} p\left(j_{k,l}^{r',r}\right) p\left(u_{k,l}^{r',r}|x_k^{r'},x_l^r,j_{k,l}^{r,r'}\right). \tag{5}$$

We let Σ_{in} and Σ_{out} to represent the covariances corresponding to inlier and outlier distributions, respectively, with $\Sigma_{in} \ll \Sigma_{out}$. The probability $p\left(u_{k,l}^{r',r}|x_k^{r'},x_l^r,j_{k,l}^{r,r'}\right)$ in Eq. (5) can be evaluated for both $j_{k,l}^{r,r'} = inlier, outlier$ using Eq. (3).



(a) Ground truth $T_r^{r'}$ (b) Inferred $T_r^{r'}$

Figure 2: Distribution of the *actual* inlier and outlier correspondences error over time evaluated using (a) ground truth and (b) inferred initial relative pose transformation $T_r^{r'}$. Using ground truth $T_r^{r'}$ the inliers can be easily distinguished from outliers (a) . In contrast, (b) shows that for arbitrary value of $T_r^{r'}$ the errors for inlier and outlier correspondences overlap, and only after estimating the transformation $T_r^{r'}$ (around pose index 140), inliers and outliers become distinguishable.

As calculating the MAP estimate over robot states $\hat{\mathcal{X}}^r = \arg\max_{\mathcal{X}^r} \sum_{\mathcal{J}^r} p\left(\mathcal{X}^r, \mathcal{J}^r | \mathcal{Z}^r\right)$ is computationally expensive, we resort to an Expectation-Maximization approach, a single iteration of which can be formulated as

$$\hat{\mathcal{X}}^r_{(i+1)} = \arg\max_{\mathcal{X}^r} p\left(J^r | \hat{\mathcal{X}}^r_{(i)}, \mathcal{Z}^r\right) \log\left[p\left(\mathcal{X}^r | \hat{\mathcal{J}}^r_{(i)}, \mathcal{Z}^r\right)\right], \quad (6)$$

where the notation (i) represents an iteration number.

Observe that when the initial relative pose $T_r^{r'}$ between two robots r and r' is unknown, performing inference over Eqs. (6) or (5) is doomed to failure: since $T_r^{r'}$ is unknown and can only be arbitrarily set, each candidate multi-robot data association $(r', k, l) \in \mathcal{F}^r$ with a corresponding constraint $u_{k,l}^{r',r}$ will typically result high errors (4) *both for inlier and outlier correspondences*. However, since $\Sigma_{in} \ll \Sigma_{out}$, the probability $p\left(u_{k,l}^{r',r} | x_k^{r'}, x_l^r, j_{k,l}^{r',r}\right)$ will be higher for $j_{k,l}^{r',r} = outlier$ than $j_{k,l}^{r',r} = inlier$, *regardless* if the correspondence $(r', k, l) \in \mathcal{F}^r$ is an inlier or outlier in practice. As a result, *all* candidate correspondences in $\mathcal{F}^r$ will be identified as outliers. It is for this reason that initial relative poses *must* be first estimated so that the error in Eq. (4) could be used to distinguish between inlier and outlier correspondences.

This observation is illustrated in Fig. 2 for the candidate correspondences between the blue and red robots shown in Fig. 1. The figure summarizes the errors (4) for all such correspondences evaluated using *ground truth* value for the initial relative pose transformation $T_r^{r'}$ (Fig. 2a) and the *inferred* transformation $\hat{T}_r^{r'}$ at each time step. As seen, when using the true transformation $T_r^{r'}$, the errors substantially differ for inliers and outliers, and can be distinguished from each other. On the other hand, evaluating Eq. (4) using the arbitrarily chosen robot reference frames, results in high errors and, more importantly, the inlier and outlier error levels *overlap each other* and therefore the inlier and outliers cannot be easily distinguished. Only after this transformation is correctly established, the errors drop and a natural segmentation into inliers and outliers arises (right area in Fig. 2b).

Consequently, we propose first to infer the transformations $T_r^{r'}$ and only then proceed to infer robot trajectories via the EM optimization (6). Our approach is based on the following *key observation* [5]: given local robot trajectories, *each* candidate multi-robot correspondence $(r', k, l) \in \mathcal{F}^r$, regardless if it is inlier or outlier, suggests a solution for the transformation $T_r^{r'}$. However, *only* the inlier correspondences will produce similar transformations, while those calculated from outlier correspondences will typically disagree amongst each other, *unless* these outliers are caused by perceptual aliasing.

Yet, real-world scenarios often exhibit some level of measurement aliasing, which typically leads to *multiple* clusters and further complicate the identification of the correct transformation $T_r^{r'}$. See Figs. 3a and 4a. Our approach addresses this challenge within a *model selection* framework, where we calculate the probability of each hypothesis $h = \{I, O\}$, representing a partition of the multi-robot correspondences $\mathcal{J}$ into inliers I and O. The set of hypotheses $\mathcal{H}$ is determined by identifying the most dominant clusters.

Furthermore, we address the question whether sufficient amount of information has been accumulated to make a decision regarding the most likely hypothesis, which is used for initializing the transformation $\hat{T}_r^{r'}$ participating in the EM optimization (6). In other words, given a set $\mathcal{H}$ there is always a hypothesis with the highest measurement likelihood (i.e. highest number of inliers); how to decide if that hypothesis is unambiguous and should be indeed chosen?

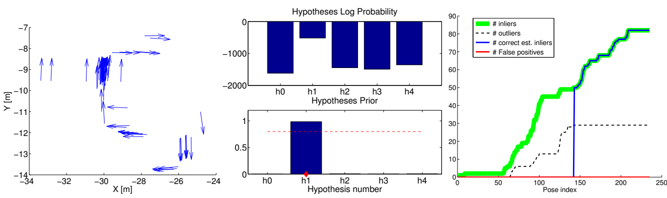
This aspect is particularly crucial in the *incremental* setting in the context of *perceptual aliasing* that can lead to a dominant hypothesis of consistent outliers (see Fig. 4a). Relying only on the measurement likelihood term, it is easy to mistakenly choose this incorrect hypothesis, which will lead to catastrophic results - see Fig. 4b where the robot trajectories are completely misaligned.

To address this crucial issue we argue the hypothesis prior, $p\left(h | \hat{\mathcal{X}}^r\right)$, can provide insight as to how likely is the hypothesis $h = \{I, O\}$ is the first place. Given the robot local trajectories, one can reason about the partitioning of the set $\mathcal{F}^r$ into inlier and outlier sets I and O, respectively, even without the actual measurements of the corresponding constraints. We introduce an additional hypothesis $h_0 = \{I_0, O_0\}$ into the set $\mathcal{H}$: this hypothesis, that we denote as the *null-hypothesis*, corresponds to perceptual aliasing, i.e. it represents the possibility that *all* of the correspondences are actually outliers ($I_0 \equiv \phi$).

The prior probabilities of all hypotheses can now be calculated. Our basic assumption is that since the robots are operating in closed indoor environments, they will eventually observe common places. Each given candidate correspondence can therefore represent the same place, observed by two different robots, or two different places. However, the number of unique places is unknown ahead of time. To capture this probabilistically, we resort to the Chinese restaurant process (see e.g. [1]) and formulate the following lemma.

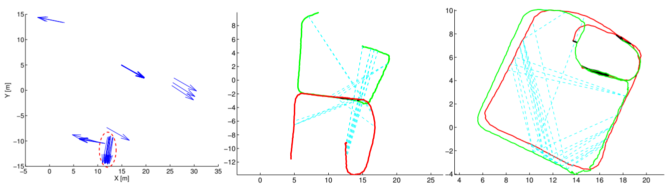
Lemma. *The* ratio *between hypotheses prior probabilities* increases *as more information is accumulated.*

The implication of this Lemma is that, as more information is accumulated, it becomes possible to disambiguate between the different hypotheses. Fig. 5 shows this process over several time instances; Fig. 4c shows the resulting correct alignment



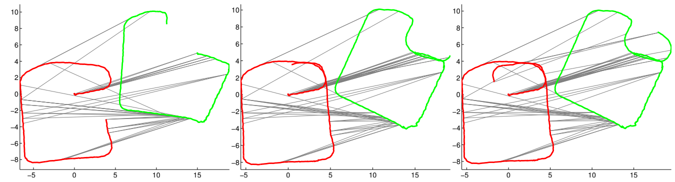
(a) (b) (c)

Figure 3: (a) Distribution of the transformations $T_r^{r'}$ calculated for each correspondence in set $\mathcal{F}^r$ for red and blue robots. (c) Hypothesis posterior (in log-space) and prior probabilities. The hypothesis h_1, corresponding to the dominant cluster in (a), is chosen. (d) Actual and inferred inliers and outliers. The method correctly identified all inliers.

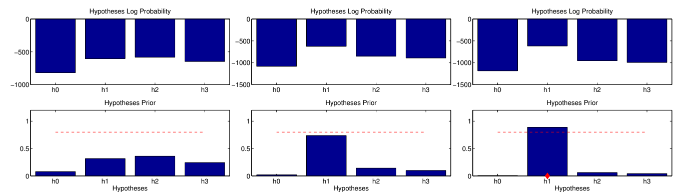


(a) (b) (c)

Figure 4: Effect of *perceptual aliasing:* (a) Distribution of the transformations $T_r^{r'}$ calculated for each correspondence in set $\mathcal{F}^r$ for red and green robots. Green robot is traveling in *opposite* direction but observes similar environments, leading to the (emphasized) cluster of *consistent* outliers. (b) Without using hypothesis prior, this cluster is chosen and the EM optimization (6) results in robot trajectories erroneously aligned. (c) Using hypothesis prior the correct hypothesis is selected (see Fig. 5), after sufficient information is accumulated and robot trajectories are then correctly aligned.



(a)



(b)

Figure 5: Establishing common reference frame between green and red robots: (a) Candidate correspondences $\mathcal{F}^r$ at three different time instances. (b) Hypothesis posterior (in log-space) and prior probabilities for three time indices. h_0 represents the null-hypothesis. Hypothesis h_1 is chosen after its prior probability crosses a predefined threshold (0.8) and because it has the highest posterior probability.

between the robots after choosing the correct hypothesis.

IV. Conclusions

We presented an approach for distributed and incremental inference by a group of collaborating robots that are initially unaware of each other's position and without assuming multi-robot data association to be given. We formulate this problem within an EM framework that, starting from promising initial guesses, converges to a number of locally-optimal hypotheses regarding data association and reference frame transformations. Choosing the correct hypothesis is challenging in the incremental setting due to perceptual aliasing and since there may be insufficient data to make this decision reliably. We address these challenges by developing a model-based selection approach for choosing the most probable hypothesis, and using the Chinese restaurant process to disambiguate the hypotheses prior probability as more information is accumulated.

References

[1] D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. In *Advances in Neural Information Processing Systems (NIPS)*, volume 16, 2003.

[2] A. Cunningham, K. Wurm, W. Burgard, and F. Dellaert. Fully distributed scalable smoothing and mapping with robust multi-robot data association. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, St. Paul, MN, 2012.

[3] A. Howard. Multi-robot simultaneous localization and mapping using particle filters. *Intl. J. of Robotics Research*, 25(12):1243–1256, 2006.

[4] V. Indelman, P. Gurfil, E. Rivlin, and H. Rotstein. Graph-based distributed cooperative navigation for a general multi-robot measurement model. *Intl. J. of Robotics Research*, 31(9), August 2012.

[5] V. Indelman, E. Nelson, N. Michael, and F. Dellaert. Multi-robot pose graph localization and data association from unknown initial relative poses via expectation maximization. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2014.

[6] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller. Multiple relative pose graphs for robust cooperative mapping. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, pages 3185–3192, Anchorage, Alaska, May 2010.

[7] Y. Latif, C. D. C. Lerma, and J. Neira. Robust loop closing over time. In *Robotics: Science and Systems (RSS)*, 2012.

[8] G. H. Lee, F. Fraundorfer, and M. Pollefeys. Robust pose-graph loop-closures with expectation-maximization. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[9] E. Montijano, R. Aragues, and C. Sagues. Distributed data association in robotic networks with cameras and limited communications. *IEEE Trans. Robotics*, 29(6):1408–1423, 2013.

[10] S.I. Roumeliotis and G.A. Bekey. Distributed multi-robot localization. *IEEE Trans. Robot. Automat.*, August 2002.

[11] Niko Sünderhauf and Peter Protzel. Switchable constraints vs. max-mixture models vs. rrr–a comparison of three approaches to robust pose graph slam. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013.